

UTILITY PATENT APPLICATION TRANSMITTAL		Attorney Docket No. 120087.405	
<div style="display: flex; align-items: center;"><div style="writing-mode: vertical-rl; transform: rotate(180deg); font-size: small; margin-right: 5px;">03/03/98</div><div style="border: 1px solid black; padding: 2px; font-size: x-small;">JC541 U.S. PTO</div></div> <p style="font-size: x-small;">(Only for nonprovisional applications under 37 CFR § 1.53(b))</p>		First Inventor or Application Identifier Dwayne Bowman	
		Title IDENTIFYING THE ITEMS MOST RELEVANT TO A CURRENT QUERY BASED ON ITEMS SELECTED IN CONNECTION WITH SIMILAR QUERIES	
		Express Mail Label No. EM150270860US	
APPLICATION ELEMENTS <small>See MPEP chapter 600 concerning utility patent application contents.</small>		ADDRESS TO: Box Patent Application Assistant Commissioner for Patents Washington, D.C. 20231	
<div style="display: flex;"><div style="flex: 1;"><p>1. <input checked="" type="checkbox"/> General Authorization Form <small>(Submit an original and a duplicate for fee processing)</small></p><p>2. <input checked="" type="checkbox"/> Specification [Total Pages] 28 <small>(preferred arrangement set forth below)</small></p><ul style="list-style-type: none">- Descriptive Title of the Invention- Cross References to Related Applications- Statement Regarding Fed sponsored R & D- Reference to Microfiche Appendix- Background of the Invention- Brief Summary of the Invention- Brief Description of the Drawings <i>(if filed)</i>- Detailed Description- Claim(s)- Abstract of the Disclosure<p>3. <input checked="" type="checkbox"/> Drawing(s) <small>(35 USC 113)</small> [Total Sheets] 9</p><p>4. Oath or Declaration [Total Pages] </p><div style="margin-left: 20px;"><p>a. <input type="checkbox"/> Newly executed (original or copy)</p><p>b. <input type="checkbox"/> Copy from a prior application (37 CFR 1.63(d)) <small>(for continuation/divisional with Box 17 completed)</small></p><div style="margin-left: 20px;"><p>i. <input type="checkbox"/> <u>DELETION OF INVENTOR(S)</u> Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR 1.63(d)(2) and 1.33(b)</p></div></div><p>5. <input type="checkbox"/> Incorporation By Reference <small>(useable if box 4b is checked)</small> The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied under Box 4b, is considered as being part of the disclosure of the accompanying application and is hereby incorporated by reference therein.</p></div><div style="flex: 1; vertical-align: top;"><p>6. <input type="checkbox"/> Microfiche Computer Program <i>(Appendix)</i></p><p>7. Nucleotide and Amino Acid Sequence Submission <small>(if applicable, all necessary)</small></p><div style="margin-left: 20px;"><p>a. <input type="checkbox"/> Computer-Readable Copy</p><p>b. <input type="checkbox"/> Paper Copy (identical to computer copy)</p><p>c. <input type="checkbox"/> Statement verifying identity of above copies</p></div></div></div>		ACCOMPANYING APPLICATION PARTS	
<p>8. <input type="checkbox"/> Assignment Papers (cover sheet & document(s))</p> <p>9. <input type="checkbox"/> 37 CFR 3.73(b) Statement <input type="checkbox"/> Power of Attorney <small>(when there is an assignee)</small></p> <p>10. <input type="checkbox"/> English Translation Document <i>(if applicable)</i></p> <p>11. <input type="checkbox"/> Information Disclosure Statement (IDS)/PTO-1449 <input type="checkbox"/> Copies of IDS Citations</p> <p>12. <input type="checkbox"/> Preliminary Amendment</p> <p>13. <input checked="" type="checkbox"/> Return Receipt Postcard</p> <p>14. <input type="checkbox"/> Small Entity Statement(s) <input type="checkbox"/> Statement filed in prior application, Status still proper and desired</p> <p>15. <input type="checkbox"/> Certified Copy of Priority Document(s) <small>(if foreign priority is claimed)</small></p> <p>16. <input checked="" type="checkbox"/> Other: <u>Certificate of Express Mail</u></p>			
<p>17. If a CONTINUING APPLICATION, check appropriate box and supply the requisite information below and in a preliminary amendment</p> <div style="display: flex; align-items: center;"><div style="margin-right: 20px;"><input type="checkbox"/> Continuation</div><div style="margin-right: 20px;"><input type="checkbox"/> Divisional</div><div style="margin-right: 20px;"><input type="checkbox"/> Continuation-In-Part (CIP)</div><div>of prior Application No.: _____</div></div> <p>Prior application information: Examiner _____ Group / Art Unit _____</p>			
CORRESPONDENCE ADDRESS			
Steven D. Lawrenz Seed and Berry LLP 6300 Columbia Center 701 Fifth Avenue Seattle, Washington 98104-7092 (206) 622-4900 phone (206) 682-6031 fax			

Respectfully submitted,

TYPED or PRINTED NAME Steven D. LawrenzREGISTRATION NO. 37,376SIGNATURE Date 2 March 1998

\\Main\wpn\120087 - Amazon.com\405\Forms\app transmittal.doc

IDENTIFYING THE ITEMS MOST RELEVANT TO A CURRENT QUERY BASED ON ITEMS SELECTED IN CONNECTION WITH SIMILAR QUERIES

TECHNICAL FIELD

The present invention is directed to the field of query processing.

5 BACKGROUND OF THE INVENTION

Many World Wide Web sites permit users to perform searches to identify a small number of interesting items among a much larger domain of items. As an example, several web index sites permit users to search for particular web sites among most of the known web sites. Similarly, many online merchants, such as
10 booksellers, permit users to search for particular products among all of the products that can be purchased from a merchant. In many cases, users perform searches in order to ultimately find a single item within an entire domain of items.

In order to perform a search, a user submits a query containing one or more query terms. The query also explicitly or implicitly identifies a domain of items
15 to search. For example, a user may submit a query to an online bookseller containing terms that the user believes are words in the title of a book. A query server program processes the query to identify within the domain items matching the terms of the query. The items identified by the query server program are collectively known as a query result. In the example, the query result is a list of books whose titles contain
20 some or all of the query terms. The query result is typically displayed to the user as a list of items. This list may be ordered in various ways. For example, the list may be ordered alphabetically or numerically based on a property of each item, such as the title, author, or release date of each book. As another example, the list may be ordered based on the extent to which each identified item matches the terms of the query.

25 When the domain for a query contains a large number of items, it is common for query results to contain tens or hundreds of items. Where the user is performing the search in order to find a single item, application of conventional approaches to ordering the query result often fail to place the sought item or items near

the top of the query result, so that the user must read through many other items in the query result before reaching the sought item. In view of this disadvantage of conventional approaches to ordering query results, a new, more effective technique for automatically ordering query results in accordance with collective and individual user behavior would have significant utility.

SUMMARY OF THE INVENTION

The present invention provides a software facility ("the facility") for identifying the items most relevant to a current query based on items selected in connection with similar queries. The facility preferably generates ranking values for items indicating their level of relevance to the current query, which specifies one or more query terms. The facility generates a ranking value for an item by combining rating scores, produced by a rating function, that each correspond to the level of relevance of the item to queries containing one of the ranking values. The rating function preferably retrieves a rating score for the combination of an item and a term from a rating table generated by the facility. The scores in the rating table preferably reflect, for a particular item and term, how often users have selected the item when the item has been identified in query results produced for queries containing particular term.

In different embodiments, the facility uses the rating scores to either generate a ranking value for each item in a query result, or generate ranking values for a smaller number of items in order to select a few items having the top ranking values. To generate a ranking value for a particular item in a query result, the facility combines the rating scores corresponding to that item and the terms of the query. In embodiments in which the goal is to generate ranking values for each item in the query result, the facility preferably loops through the items in the query results and, for each item, combines all of the rating scores corresponding to that item and any of the terms in the query. On the other hand, in embodiments in which the goal is to select a few items in the query result having the largest ranking values, the facility preferably loops through the terms in the query, and, for each item, identifies the top few rating scores for that

term and any item. The facility then combines the scores identified for each item to generate ranking values for a relatively small number of items, which may include items not identified in the query result. The facility preferably orders the items of the query result in decreasing order of ranking value. The facility may also use the ranking values to subset the items in the query result to a smaller number of items.

By ordering and/or subsetting the items in the query result in this way in accordance with collective and individual user behavior rather than in accordance with attributes of the items, the facility substantially increases the likelihood that the user will quickly find within the query result the particular item or items that he or she seeks. For example, while a query result for a query containing the query terms “human” and “dynamic” may contain a book about human dynamics and a book about the effects on human beings of particle dynamics, selections by users from early query results produced for queries containing the term “human” show that these users select the human dynamics book much more frequently than they select the particle dynamics book. The facility therefore ranks the human dynamics book higher than the particle dynamics book, allowing users that are more interested in the human dynamics book to select it more easily. This benefit of the facility is especially useful in conjunction with the large, heterogeneous query results that are typically generated for single-term queries, which are commonly submitted by users.

Various embodiments of the invention base rating scores on different kinds of selection actions performed by the users on items identified in query results. These include whether the user displayed additional information about an item, how much time the user spent viewing the additional information about the item, how many hyperlinks the user followed within the additional information about the item, whether the user added the item to his or her shopping basket, and whether the user ultimately purchased the item. Embodiments of the invention also consider selection actions not relating to query results, such as typing an item’s item identifier rather than choosing the item from a query result. Additional embodiments of the invention incorporate into the ranking process information about the user submitting the query by maintaining and applying separate rating scores for users in different demographic groups, such as those

of the same sex, age, income, or geographic category. Certain embodiments also incorporate behavioral information about specific users. Further, rating scores may be produced by a rating function that combines different types of information reflecting collective and individual user preferences. Some embodiments of the invention utilize
 5 specialized strategies for incorporating into the rating scores information about queries submitted in different time frames.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a high-level block diagram showing the computer system upon which the facility preferably executes.

10 Figure 2 is a flow diagram showing the steps preferably performed by the facility in order to generate a new rating table. In step 201, the facility initializes a rating table.

Figures 3 and 4 are table diagrams showing augmentation of an item rating table in accordance with step 206 (Figure 2).

15 Figure 5 is a table diagram showing the generation of rating tables for composite periods of time from rating tables for constituent periods of time.

Figure 6 is a table diagram showing a rating table for a composite period.

Figure 7 is a flow diagram showing the steps preferably performed by the facility in order to identify user selections within a web server log.

20 Figure 8 is a flow diagram showing the steps preferably performed by the facility to order a query result using a rating table by generating a ranking value for each item in the query result.

Figure 9 is a flow diagram showing the steps preferably performed by the facility to select a few items in a query result having the highest ranking values
 25 using a rating table.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a software facility ("the facility") for identifying the items most relevant to a current query based on items selected in

connection with similar queries. The facility preferably generates ranking values for items indicating their level of relevance to the current query, which specifies one or more query terms. The facility generates a ranking value for an item by combining rating scores, produced by a rating function, that each correspond to the level of relevance of the item to queries containing one of the ranking values. The rating function preferably retrieves a rating score for the combination of an item and a term from a rating table generated by the facility. The scores in the rating table preferably reflect, for a particular item and term, how often users have selected the item when the item has been identified in query results produced for queries containing the term.

10 In different embodiments, the facility uses the rating scores to either generate a ranking value for each item in a query result, or generate ranking values for a smaller number of items in order to select a few items having the top ranking values. To generate a ranking value for a particular item in a query result, the facility combines the rating scores corresponding to that item and the terms of the query. In embodiments
15 in which the goal is to generate ranking values for each item in the query result, the facility preferably loops through the items in the query results and, for each item, combines all of the rating scores corresponding to that item and any of the terms in the query. On the other hand, in embodiments in which the goal is to select a few items in the query result having the largest ranking values, the facility preferably loops through
20 the terms in the query, and, for each item, identifies the top few rating scores for that term and any item. The facility then combines the scores identified for each item to generate ranking values for a relatively small number of items, which may include items not identified in the query result. The facility preferably orders the items of the query result in decreasing order of ranking value. The facility may also use the ranking values
25 to subset the items in the query result to a smaller number of items.

By ordering and/or subsetting the items in the query result in this way in accordance with collective and individual user behavior rather than in accordance with attributes of the items, the facility substantially increases the likelihood that the user will quickly find within the query result the particular item or items that he or she seeks.
30 For example, while a query result for a query containing the query terms “human” and

“dynamic” may contain a book about human dynamics and a book about the effects on human beings of particle dynamics, selections by users from early query results produced for queries containing the term “human” show that these users select the human dynamics book much more frequently than they select the particle dynamics book. The facility therefore ranks the human dynamics book higher than the particle dynamics book, allowing users, most of whom are more interested in the human dynamics book, to select it more easily. This benefit of the facility is especially useful in conjunction with the large, heterogeneous query results that are typically generated for single-term queries, which are commonly submitted by users.

Various embodiments of the invention base rating scores on different kinds of selection actions performed by the users on items identified in query results. These include whether the user displayed additional information about an item, how much time the user spent viewing the additional information about the item, how many hyperlinks the user followed within the additional information about the item, whether the user added the item to his or her shopping basket, and whether the user ultimately purchased the item. Embodiments of the invention also consider selection actions not relating to query results, such as typing an item’s item identifier rather than choosing the item from a query result. Additional embodiments of the invention incorporate into the ranking process information about the user submitting the query by maintaining and applying separate rating scores for users in different demographic groups, such as those of the same sex, age, income, or geographic category. Certain embodiments also incorporate behavioral information about specific users. Further, rating scores may be produced by a rating function that combines different types of information reflecting collective and individual user preferences. Some embodiments of the invention utilize specialized strategies for incorporating into the rating scores information about queries submitted in different time frames.

Figure 1 is a high-level block diagram showing the computer system upon which the facility preferably executes. As shown in Figure 1, the computer system 100 comprises a central processing unit (CPU) 110, input/output devices 120, and a computer memory (memory) 130. Among the input/output devices is a storage

device 121, such as a hard disk drive; a computer-readable media drive 122, which can be used to install software products, including the facility, which are provided on a computer-readable medium, such as a CD-ROM; and a network connection 123 for connection the computer system 100 to other computer systems (not shown). The memory 130 preferably contains a query server 131 for generating query results from queries, a query result ranking facility 132 for automatically ranking the items in a query result in accordance with collective user preferences, and item rating tables 133 used by the facility. While the facility is preferably implemented on a computer system configured as described above, those skilled in the art will recognize that it may also be implemented on computer systems having different configurations.

The facility preferably generates a new rating table periodically, and, when a query result is received, uses the last-generated rating table to rank the items in the query result. Figure 2 is a flow diagram showing the steps preferably performed by the facility in order to generate a new rating table. In step 201, the facility initializes a rating table for holding entries each indicating the rating score for a particular combination of a query term and an item identifier. The rating table preferably has no entries when it is initialized. In step 202, the facility identifies all of the query result item selections made by users during the period of time for which the rating table is being generated. The rating table may be generated for the queries occurring during a period of time such as a day, a week, or month. This group of queries is termed a “rating set” of queries. The facility also identifies the terms of the queries that produced these query results in step 202. Performance of step 202 is discussed in greater detail below in conjunction with Figure 7. In steps 204-208, the facility loops through each item selection from a query result that was made by a user during the time period. In step 204, the facility identifies the terms used in the query that produced the query result in which the item selection took place. In steps 205-207, the facility loops through each term in the query. In step 206, the facility increases the rating score in the rating table corresponding to the current term and item. Where an entry does not yet exist in the rating table for the term and item, the facility adds a new entry to the rating table for the term and item. Increasing the rating score preferably involves adding an increment

value, such as 1, to the existing rating score for the term and item. In step 207, if additional terms remain to be processed, the facility loops back to step 205 to process the next term in the query, else the facility continues in step 208. In step 208, if additional item selections remain to be processed, then the facility loops back to step 5 203 to process the next item selection, else these steps conclude.

Figures 3 and 4 are table diagrams showing augmentation of an item rating table in accordance with step 206 (Figure 2). Figure 3 shows the state of the item rating table before its augmentation. It can be seen that the table 300 contains a number of entries, including entries 301-306. Each entry contains the rating score for a particular combination of a query term and an item identifier. For example, entry 302 10 identifies the score “22” for the term “dynamics” the item identifier “1883823064”. It can be seen by examining entries 301-303 that, in query results produced from queries including the term “dynamics”, the item having item identifier “1883823064” has been selected by users more frequently than the item having item identifier “9676530409”, 15 and much more frequently than the item having item identifier “0801062272”. In additional embodiments, the facility uses various other data structures to store the rating scores, such as sparse arrays.

In augmenting the item rating table 300, the facility identifies the selection of the item having item identifier “1883823064” from a query result produced 20 by a query specifying the query terms “human” and “dynamics”. Figure 4 shows the state of the item rating table after the item rating table is augmented by the facility to reflect this selection. It can be seen by comparing entry 405 in item rating table 400 to entry 305 in item rating table 300 that the facility has incremented the score for this entry from “45” to “46”. Similarly, the facility has incremented the rating score for this 25 item identifier the term “dynamics” from “22” to “23”. The facility augments the rating table in a similar manner for the other selections from query results that it identifies during the time period.

Rather than generating a new rating table from scratch using the steps shown in Figure 2 each time new selection information becomes available, the facility 30 preferably generates and maintains separate rating tables for different constituent time

periods, of a relatively short length, such as one day. Each time a rating table is generated for a new constituent time period, the facility preferably combines this new rating table with existing rating tables for earlier constituent time periods to form a rating table for a longer composite period of time. Figure 5 is a table diagram showing the generation of rating tables for composite periods of time from rating tables for constituent periods of time. It can be seen in Figure 5 that rating tables 501-506 each correspond to a single day between 8-Feb-98 and 13-Feb-98. Each time a new constituent period is completed, the facility generates a new rating table reflecting the user selections made during that constituent period. For example, at the end of 12-Feb-98, the facility generates rating table 505, which reflects all of the user selections occurring during 12-Feb-98. After the facility generates a new rating table for a completed constituent period, the facility also generates a new rating table for a composite period ending with that constituent period. For example, after generating the rating table 505 for the constituent period 12-Feb-98, the facility generates rating table 515 for the composite period 8-Feb-98 to 12-Feb-98. The facility preferably generates such a rating table for a composite period by combining the entries of the rating tables for the constituent periods making up the composite period, and combining the scores of corresponding entries, for example, by summing them. In one preferred embodiment, the scores and rating tables for more recent constituent periods are weighted more heavily than those in rating tables for less recent constituent periods. When ranking query results, the rating table for the most recent composite period is preferably used. That is, until rating table 516 can be generated, the facility preferably uses rating table 515 to rank query results. After rating table 516 is generated, the facility preferably uses rating table 516 to rank query results. The lengths of both constituent periods and composite periods are preferably configurable.

Figure 6 is a table diagram showing a rating table for a composite period. By comparing the item rating table 600 shown in Figure 6 to item rating table 400 shown in Figure 4, it can be seen that the contents of rating table 600 constitute the combination of the contents of rating table 400 with several other rating tables for constituent periods. For example, the score for entry 602 is "116", or about five times

the score for corresponding entry 402. Further, although rating table 400 does not contain an entry for the term “dynamics” and the item identifier “1887650024”, entry 607 has been added to table 600 for this combination of term and item identifier, as a corresponding entry occurs in a rating table for one of the other constituent periods within the composite period.

The process used by the facility to identify user selections is dependent upon both the kind of selection action used by the facility and the manner in which the data relating to such selection actions is stored. One preferred embodiment uses as its selection action requests to display more information about items identified in query results. In this embodiment, the facility extracts this information from logs generated by a web server that generates query results for a user using a web client, and allows the user to select an item with the web client in order display additional information about it. A web server generally maintains a log detailing of all the HTTP requests that it has received from web clients and responded to. Such a log is generally made up of entries, each containing information about a different HTTP request. Such logs are generally organized chronologically. Log Entry 1 below is a sample log entry showing an HTTP request submitted by a web client on behalf of the user that submits a query.

```

1. Friday, 13-Feb-98 16:59:27
2. User Identifier=82707238671
3. HTTP_REFERER=http://www.amazon.com/book_query_page
4. PATH_INFO=/book_query
5. author="Seagal"
6. title="Human Dynamics"

```

Log Entry 1

It can be seen by the occurrence of the keyword “book_query” in the “PATH_INFO” line 4 of Log Entry 1 that this log entry corresponds to a user’s submission of a query. It further can be seen in term lines 5 and 6 that the query includes the terms “Seagal”, “Human”, and “Dynamics”. In line 2, the entry further contains a user identifier corresponding to the identity of the user and, in some embodiments, also to this particular interaction with the web server.

In response to receiving the HTTP request documented in Log Entry 1, the query server generates a query result for the query and returns it to the web client submitting the query. Later the user selects an item identified in the query result, and the web client submits another HTTP request to display detailed information about the selected item. Log Entry 2, which occurs at a point after Log Entry 1 in the log, describes this second HTTP request.

1. Friday, 13-Feb-98 17:02:39
2. User Identifier=82707238671
- 10 3. HTTP_REFERER=http://www.amazon.com/book_query
4. PATH_INFO=/ISBN=1883823064

Log Entry 2

- 15 By comparing the user identifier in line 2 of Log Entry 2 to the user identifier in line 2 of Log Entry 1, it can be seen that these log entries correspond to the same user and time frame. In the "PATH_INFO" line 4 of Log Entry 2, it can be seen that the user has selected an item having item identifier ("ISBN") "1883823064". It can further be seen from the occurrence of the keyword "book_query" on the "HTTP_REFERER" line 3
- 20 that the selection of this item was from a query result.

Where information about user selections is stored in web server logs such as those discussed above, the facility preferably identifies user selections by traversing these logs. Such traversal can occur either in a batch processing mode after a log for a specific period of time has been completely generated, or in a real-time processing mode so that log entries are processed as soon as they are generated.

Figure 7 is a flow diagram showing the steps preferably performed by the facility in order to identify user selections within a web server log. In step 701, the facility positions a first pointer at the top, or beginning, of the log. The facility then repeats steps 702-708 until the first pointer reaches the end of the log. In step 703, the facility traverses forward with the first pointer to the next item selection event. In terms of the log entry shown above, step 703 involves traversing forward through log entries until one is found that contains in its "HTTP_REFERER" line a keyword denoting a search entry, such as "book_query". In step 704, the facility extracts from this item

selection event the identity of the item that was selected and session identifier that identifies the user that selected the item. In terms of the log entries above, this involves reading the ten-digit number following the string “ISBN=” in the “PATH_INFO” line of the log entry, and reading the user identifier from the “User Identifier” line of the log entry. Thus, in Log Entry 2, the facility extracts item identifier “1883823064” and session identifier “82707238761”. In step 705, the facility synchronizes the position of the second pointer with the position of the first pointer. That is, the facility makes the second pointer point to the same log entry as the first pointer. In step 706, the facility traverses backwards with the second pointer to a query event having a matching user identifier. In terms of the log entries above, the facility traverses backward to the log entry having the keyword “book_query” in its “PATH_INFO” line, and having a matching user identifier on its “User Identifier” line. In step 707, the facility extracts from the query event to which the second pointer points the terms of the query. In terms of the query log entries above, the facility extracts the quoted words from the query log entry to which the second pointer points, in the lines after the “PATH_INFO” line. Thus, in Log Entry 1, the facility extracts the terms “Seagal”, “Human”, and “Dynamics”. In step 708, if the first pointer has not yet reached the end of the log, then the facility loops back to step 702 to continue processing the log, else these steps conclude.

When other selection actions are used by the facility, extracting information about the selection from the web server log can be somewhat more involved. For example, where the facility uses purchase of the item as the selection action, instead of identifying a log entry describing a request by the user for more information about an item, like Log Entry 1, the facility instead identifies a log entry describing a request to purchase items in a “shopping basket.” The facility then traverses backwards in the log, using the entries describing requests to add items to and remove items from the shopping basket to determine which items were in the shopping basket at the time of the request to purchase. The facility then continues traversing backward in the log to identify the log entry describing the query, like Log Entry 2, and to extract the search terms.

Rather than relying solely on a web server log where item purchase is the selection action that is used by the facility, the facility alternatively uses a database separate from the web server log to determine which items are purchased in each purchase transaction. This information from the database is then matched up with the log entry containing the query terms for the query from which item is selected for purchase. This hybrid approach, using the web server logs and a separate database, may be used for any of the different kinds of selection actions. Additionally, where a database separate from the web server log contains all the information necessary to augment the rating table, the facility may use the database exclusively, and avoid traversing the web server log.

The facility uses rating tables that it has generated to generate ranking values for items in new query results. Figure 8 is a flow diagram showing the steps preferably performed by the facility to order a query result using a rating table by generating a ranking value for each item in the query result. In steps 801-807, the facility loops through each item identified in the query result. In step 802, the facility initializes a ranking value for the current item. In steps 803-805, the facility loops through each term occurring in the query. In step 804, the facility determines the rating score contained by the most recently-generated rating table for the current term and item. In step 805, if any terms of the query remain to be processed, then the facility loops up to step 803, else the facility continues in step 806. In step 806, the facility combines the scores for the current item to generate a ranking value for the item. As an example, with reference to Figure 6, in processing datum having item identifier “1883823064”, the facility combines the score “116” extracted from entry 602 for this item and the term “dynamics”, and the score “211” extracted from entry 605 for this item and the term “human”. Step 806 preferably involves summing these scores. These scores may be combined in other ways, however. In particular, scores may be adjusted to more directly reflect the number of query terms that are matched by the item, so that items that match more query terms than others are favored in the ranking. In step 807, if any items remain to be processed, the facility loops back to step 801 to process the next item, else the facility continues in step 808. In step 808, the facility displays the

items identified in the query result in accordance with the ranking values generated for the items in step 806. Step 808 preferably involves sorting the items in the query result in decreasing order of their ranking values, and/or subsetting the items in the query result to include only those items above a threshold ranking value, or only a predetermined number of items having the highest ranking values. After step 808, these steps conclude.

Figure 9 is a flow diagram showing the steps preferably performed by the facility to select a few items in a query result having the highest ranking values using a rating table. In steps 901-903, the facility loops through each term in the query.

10 In step 902, the facility identifies among the table entries for the current term and those entries having the three highest rating scores. For example, with reference to Figure 6, if the only entries in item rating table 600 for the term “dynamics” are entries 601, 602, 603, and 607, the facility would identify entries 601, 602, and 603, which are the entries for the term “dynamics” having the three highest rating scores. In additional preferred

15 embodiments, a small number of table entries other than three is used. In step 903, if additional terms remain in the query to be processed, then the facility loops back to step 901 to process the next term in the query, else the facility continues in step 904. In steps 904-906, the facility loops through each unique item among the identified entries. In step 905, the facility combines all of the scores for the item among the identified

20 entries. In step 906, if additional unique items remain among the identified entries to be processed, then the facility loops back to step 904 to process the next unique item, else the facility continues in step 907. As an example, if, in item rating table 600, the facility selected entries 601, 602, and 603 for the term “dynamics”, and selected entries 604, 605, and 606 for the term “human”, then the facility would combine the scores

25 “116” and “211” for the item having item identifier “1883823064”, and would use the following single scores for the remaining item identifiers: “77” for the item having item identifier “0814403484”, “45” for the item having item identifier “9676530409”, “12” for the item having item identifier “6303702473”, and “4” for the item having item identifier “0801062272”. In step 907, the facility selects for prominent display items

30 having the top three combined scores. Because the facility in step 907 selects items

without regard for their presence in the query result, the facility may select items that are not in the query result. This aspect of this embodiment is particularly advantageous in situations in which a complete query result is not available when the facility is invoked. Such as the case, for instance, where the query server only provides a portion
5 of the items satisfying the query at a time. In additional embodiments, the facility selects a small number of items having the top combined scores that is other than three. In the example discussed above, the facility would select for prominent display the items having item identifiers "1883823064", "0814403484", and "9676530409". After step 907, these steps conclude.

10 While the present invention has been shown and described with reference to preferred embodiments, it will be understood by those skilled in the art that various changes or modifications in form and detail may be made without departing from the scope of the invention. For example, the facility may be used to rank query results of all types. The facility may use various formulae to determine in the case of
15 each item selection, the amount by which to augment rating scores with respect to the selection. Further, the facility may employ various formulae to combine rating scores into a ranking value for an item. The facility may also use a variety of different kinds of selection actions to augment the rating table, and may augment the rating table for more than one kind of selection action at a time. Additionally, the facility, may
20 augment the rating table to reflect selections by users other than human users, such as software agents or other types of artificial users.

CLAIMS

We claim:

1 1. A method in a computer system for ranking items in a search result, the
 2 method comprising the steps of:
 3 receiving a rating set of queries, each query in the rating set specifying one or
 4 more terms;
 5 for each query in the rating set,
 6 generating a query result identifying one or more items satisfying the
 7 query;
 8 allowing a user to select one or more of the items identified in the
 9 query result;
 10 for each item selected from the query result, for each term specified by
 11 the query, increasing a rating value corresponding to the combination of the selected item and
 12 the term specified by the query, the rating value indicating a relative frequency with which
 13 users have selected the selected item when the selected item has been identified in query
 14 results generated from queries containing the term specified by the query;
 15 receiving a distinguished query specifying one or more terms;
 16 generating a distinguished query result identifying a plurality of items
 17 satisfying the distinguished query; and
 18 for each item identified in the distinguished query result,
 19 for the rating values corresponding to the combination of the item
 20 identified in the distinguished query result and one of the terms specified by the distinguished
 21 query, combining these rating values to generate a ranking value for the item within the
 22 distinguished query result.

1 2. The method of claim 1, further comprising the step of imposing on the
 2 items identified in the distinguished query result an order in which the ranking values of the
 3 items monotonically decreases.

1 3. The method of claim 1, further comprising the step of using the
2 ranking values of the items identified in the distinguished query result to create a proper
3 subset of the items.

1 4. The method of claim 3 wherein the creating step creates a subset of the
2 items identified in the distinguished query result that contains all of the items whose ranking
3 values exceed a predetermined minimum ranking value.

1 5. The method of claim 3 wherein the creating step creates a subset of the
2 items identified in the distinguished query result that contains all of the items whose ranking
3 values exceed a predetermined minimum ranking value.

1 6. The method of claim 1 wherein the increasing step increases rating
2 values for selections made to display additional information about items.

1 7. The method of claim 1 wherein the increasing step increases rating
2 values for selections made to purchase items.

1 8. The method of claim 1 wherein the increasing step increases rating
2 values for selections made to add items to a tentative list of purchases.

1 9. The method of claim 1 wherein the increasing step increases rating
2 values for selections of portions of detailed information displayed about items.

1 10. The method of claim 1 wherein the increasing step increases rating
2 values for units of time for which the user displays detailed information about items.

1 11. A method in a computer system for ranking items in a search result, the
2 method comprising the steps of:

3 for each of a multiplicity of search terms, compiling data indicating the extent
4 to which users have selected each of a multiplicity of items when returned in search results
5 produced from queries containing the search term;

6 receiving a query and a search result, the received query containing a term
7 among the multiplicity of terms, the received search result identifying a plurality of items
8 among the multiplicity of items that satisfy the received query; and

9 using the compiled data to rank at least a portion of the items identified in the
10 received search result in accordance with the extent to which users have selected each of the
11 plurality of items identified in the received search result when returned in search results
12 produced from queries containing the search term contained in the received query.

1 12. The method of claim 11 wherein at least a portion of the users are
2 identified with one of a plurality of demographic groups, and wherein the compiling step
3 compiles, for each demographic group of the plurality of demographic groups, data indicating
4 the extent to which users identified with the demographic group have selected each of a
5 multiplicity of items when returned in search results produced from queries containing the
6 search term, and wherein the received query is submitted on behalf of a distinguished user
7 identified with a distinguished demographic group, and wherein the ranking step uses the
8 compiled data to rank the items identified in the received search result in accordance with the
9 extent to which users identified with the distinguished demographic group have selected each
10 of the plurality of items identified in the received search result when returned in search results
11 produced from queries containing the search term contained in the received query.

1 13. The method of claim 11, further comprising the step of imposing on
2 the items identified in the distinguished search result an order in which the ranking values of
3 the items monotonically decreases.

1 14. The method of claim 11, further comprising the step of using the
2 ranking values of the items identified in the distinguished search result to create a proper
3 subset of the items.

1 15. The method of claim 14 wherein the creating step creates a subset of
2 the items identified in the distinguished search result that contains all of the items whose
3 ranking values exceed a predetermined minimum ranking value.

1 16. The method of claim 14 wherein the creating step creates a subset of
2 the items identified in the distinguished search result that contains all of the items whose
3 ranking values exceed a predetermined minimum ranking value.

1 17. The method of claim 11 wherein the increasing step increases rating
2 values for selections made to display additional information about items.

1 18. The method of claim 11 wherein the increasing step increases rating
2 values for selections made to purchase items.

1 19. The method of claim 11 wherein the increasing step increases rating
2 values for selections made to add items to a tentative list of purchases.

1 20. The method of claim 11 wherein the increasing step increases rating
2 values for selections of portions of detailed information displayed about items.

1 21. The method of claim 11 wherein the increasing step increases rating
2 values for units of time for which the user displays detailed information about items.

1 22. A computer-readable medium whose contents cause a computer system
2 to rank items in a search result by performing the steps of:
3 for each of a multiplicity of search terms, compiling data indicating the extent
4 to which users have selected each of a multiplicity of items when returned in search results
5 produced from queries containing the search term;

6 receiving a query and a search result, the received query containing a term
7 among the multiplicity of terms, the received search result identifying a plurality of items
8 among the multiplicity of items that satisfy the received query; and

9 using the compiled data to rank at least a portion of the items identified in the
10 received search result in accordance with the extent to which users have selected each of the
11 plurality of items identified in the received search result when returned in search results
12 produced from queries containing the search term contained in the received query.

1 23. A method in a computer system for ranking items in a search result, the
2 method comprising the steps of:

3 receiving a query specifying one or more terms;

4 generating a query result identifying a plurality of items satisfying the query;

5 and

6 for a plurality of items identified in the query result, combining ratings of
7 frequencies with which users selected the item in earlier queries specifying one or more terms
8 of the query to producing a ranking value for the item.

1 24. The method of claim 23, further comprising the step of adjusting the
2 produced ranking values to reflect the number of terms specified by the query that are
3 matched by the item.

1 25. The method of claim 23, further comprising the step of imposing on
2 the items identified in the distinguished query result an order in which the ranking values of
3 the items monotonically decreases.

1 26. The method of claim 23, further comprising the step of using the
2 ranking values of the items identified in the distinguished query result to create a proper
3 subset of the items.

1 27. The method of claim 26 wherein the creating step creates a subset of
2 the items identified in the distinguished query result that contains all of the items whose
3 ranking values exceed a predetermined minimum ranking value.

1 28. The method of claim 26 wherein the creating step creates a subset of
2 the items identified in the distinguished query result that contains all of the items whose
3 ranking values exceed a predetermined minimum ranking value.

1 29. A computer-readable medium whose contents cause a computer system
2 to rank items in a search result by performing the steps of:
3 receiving a query specifying one or more terms;
4 generating a query result identifying a plurality of items satisfying the query;
5 and
6 for each item identified in the query result, combining the relative frequencies
7 with which users selected the item in earlier queries specifying each of the terms of the query
8 to producing a ranking value for the item.

1 30. The computer-readable medium of claim 29 wherein the contents of
2 the computer-readable medium further cause the computer system to perform the step of
3 adjusting the ranking value produced for each item identified in the query result to reflect the
4 number of terms specified by the query that are matched by the item.

1 31. A method in a computer system for identifying significant items
2 satisfying a query, the method comprising the steps of:
3 receiving a query specifying a plurality of terms;
4 for each term specified by the query, identifying a predetermined number of
5 items that users have selected most frequently in earlier queries specifying the term;
6 for each unique item identified in the identifying step, producing a ranking
7 value for the item by combining one or more indications of the relative frequencies with

8 which users have selected the item in earlier queries specifying one of the terms specified by
 9 the received query; and

10 selecting as the significant items a predetermined number of items having the
 11 largest ranking values.

1 32. The method of claim 31, further comprising the step of displaying
 2 information relating to the selected items in a manner distinct from that in which information
 3 about unselected items is displayed.

1 33. A computer-readable medium whose contents cause a computer system
 2 to identify the most significant items satisfying a query by performing the steps of:
 3 receiving a query specifying one or more terms;
 4 for each term specified by the query, identifying a predetermined number of
 5 items having the largest relative frequencies with which users selected the items in earlier
 6 queries specifying the term;
 7 for each unique item identified in the identifying step, combining the relative
 8 frequencies for the item to produce a ranking value for the item; and
 9 selecting as the most significant items a predetermined number of items
 10 having the largest ranking values.

1 34. A computer system for ranking items in a search result, comprising:
 2 a query memory that stores information about previously submitted queries
 3 and items selected from the query results of previously submitted queries;
 4 a query receiver that receives queries each specifying one or more terms;
 5 a query server that generating a query result for each query received by the
 6 query receiver that identifies a plurality of items satisfying the query; and
 7 an item ranking subsystem that, for each query result generated by the query
 8 server, for at least a portion of the items identified in the query result, combines from the
 9 contents of the query memory the relative frequencies with which users selected the item in

10 earlier queries specifying each of the terms of the query to producing a ranking value for the
11 item.

1 35. A computer memory containing a ranked search result items data
2 structure for a distinguished search comprising a plurality of entries, each entry identifying an
3 item satisfying the distinguished search, each entry further specifying a quantitative ranking
4 of the relevance of the identified item, the quantitative ranking reflecting the extent to which
5 users have selected the identified item when it has been among the items satisfying foregoing
6 searches similar to the distinguished search.

1 36. A method in a computer system for compiling statistics usable to rank
2 items in a distinguished query result produced for a distinguished query, the method
3 comprising the steps of:
4 receiving a rating set of queries, each query in the rating set specifying one or
5 more terms;
6 for each query in the rating set,
7 generating a query result identifying one or more items satisfying the
8 query;
9 allowing a user to select one or more of the items identified in the
10 query result; and
11 for items selected from the query result, for terms specified by the
12 query, adjusting a rating score corresponding to the combination of the selected item and the
13 term specified by the query, the rating score indicating the relative frequency with which
14 users have selected the selected item when the selected item has been identified in search
15 results generated from queries containing the search term specified by the query, to produce
16 rating scores usable to rank items in a distinguished query result produced for a distinguished
17 query.

1 37. The method of claim 36 wherein the adjusting step is performed in
2 real-time with user selections.

1 38. The method of claim 36 wherein the adjusting step is performed
2 periodically for a batch of earlier-occurring user selections.

1 39. The method of claim 36 wherein the queries received in the receiving
2 step are each identified with a different time period, and wherein the adjusting step for each
3 query increases a rating score corresponding to the time period with which the query is
4 identified, to produce a different body of rating scores for each time period, and wherein the
5 method further includes the step of combining a subset of the produced bodies of rating
6 scores to generate a composite body of rating scores usable to order items in a distinguished
7 query result produced for a distinguished query.

1 40. The method of claim 39 wherein the combining step weights scores
2 from different bodies of rating scores differently depending on the recency of the time periods
3 for which the bodies of rating scores are produced.

1 41. The method of claim 36 wherein each query term is specified for one
2 of a plurality of query fields, and wherein the adjusting step for each term increases a rating
3 score corresponding to the query field for which the term is identified, to produce different
4 rating scores for each different query field for which a query term is specified.

1 42. The method of claim 36, further comprising the steps of:
2 for each received query, generating a query log entry, the query log entry
3 containing the terms specified by the query and a query identifier identifying the query;
4 for each user selection of an item identified in a query result generated for a
5 received query, generating a selection log entry, the selection log entry containing a query
6 identifier identifying the query and an item identifier identifying the item selected; and
7 identifying each user selection by:
8 identifying a selection log entry;

9 extracting from the identified selection log entry the query identifier
 10 and the item identifier;
 11 identifying a query log entry containing the extracted query identifier;
 12 and
 13 extracting from the identified query log entry the terms specified by the
 14 query.

1 43. The method of claim 36, further comprising the steps of:
 2 for each user selection of an item identified in a query result generated for a
 3 received query, generating a database record identifying both the terms of the query and the
 4 item selected; and
 5 identifying each user selection by:
 6 retrieving a database record, and
 7 extracting from the retrieved database record the terms of the query and
 8 the item selected.

1 44. The method of claim 36 wherein the adjusting step increases rating
 2 values for selections made to display additional information about items.

1 45. The method of claim 36 wherein the adjusting step increases rating
 2 values for selections made to purchase items.

1 46. The method of claim 36 wherein the adjusting step increases rating
 2 values for selections made to add items to a tentative list of purchases.

1 47. The method of claim 36 wherein the adjusting step increases rating
 2 values for selections of portions of detailed information displayed about items.

1 48. The method of claim 36 wherein the adjusting step increases rating
 2 values for units of time for which the user displays detailed information about items.

1 49. A computer-readable medium whose contents cause a computer system
2 to compile statistics usable to rank items in a distinguished query result produced for a
3 distinguished query by performing the steps of:

4 receiving a rating set of queries, each query in the rating set specifying one or
5 more terms;

6 for each query in the rating set,

7 generating a query result identifying one or more items satisfying the
8 query;

9 allowing a user to select one or more of the items identified in the
10 query result; and

11 for each item selected from the query result, for each term specified by
12 the query, increasing a rating score corresponding to the combination of the selected item and
13 the term specified by the query, the rating score indicating the relative frequency with which
14 users have selected the selected item when the selected item has been identified in search
15 results generated from queries containing the search term specified by the query, to produce a
16 body of rating scores usable to rank items in a distinguished query result produced for a
17 distinguished query.

1 50. The computer-readable medium of claim 31 wherein the queries
2 received in the receiving step are each identified with a different time period, and wherein the
3 increasing step for each query increases a rating score corresponding to the time period with
4 which the query is identified, to produce a different body of rating scores for each time
5 period, and wherein the computer-readable medium further causes the computer system to
6 perform the step of combining a subset of the produced bodies of rating scores to generate a
7 composite body of rating scores usable to order items in a distinguished query result produced
8 for a distinguished query.

1 51. A computer memory containing a user behavior data structure usable to
2 rank the relevance of items in a query result, the data structure comprising a plurality of rating

3 scores, each rating score corresponding both to a query term and to an item, and reflecting
4 quantitatively the extent to which users have selected the item from query results generated
5 from queries specifying the query term, such that the data structure may be used to rank items
6 in a distinguished query result produced for a distinguished query by, for each item in the
7 distinguished query result, retrieving from the data structure the rating scores corresponding
8 to the item and any term specified in the distinguished query and combining the retrieved
9 rating scores to generate a ranking value for the item.

IDENTIFYING THE ITEMS MOST RELEVANT TO A CURRENT QUERY BASED
ON ITEMS SELECTED IN CONNECTION WITH SIMILAR QUERIES

ABSTRACT OF THE DISCLOSURE

The present invention provides a software facility for identifying the items most relevant to a current query based on items selected in connection with similar queries. In preferred embodiments of the invention, the facility receives a query specifying one or more query terms. In response, the facility generates a query result identifying a plurality of items that satisfy the query. The facility then produces a ranking value for at least a portion of the items identified in the query result by combining the relative frequencies with which users selected that item from the query results generated from queries specifying each of the terms specified by the query. The facility identifies as most relevant those items having the highest ranking values.

client computer system 100

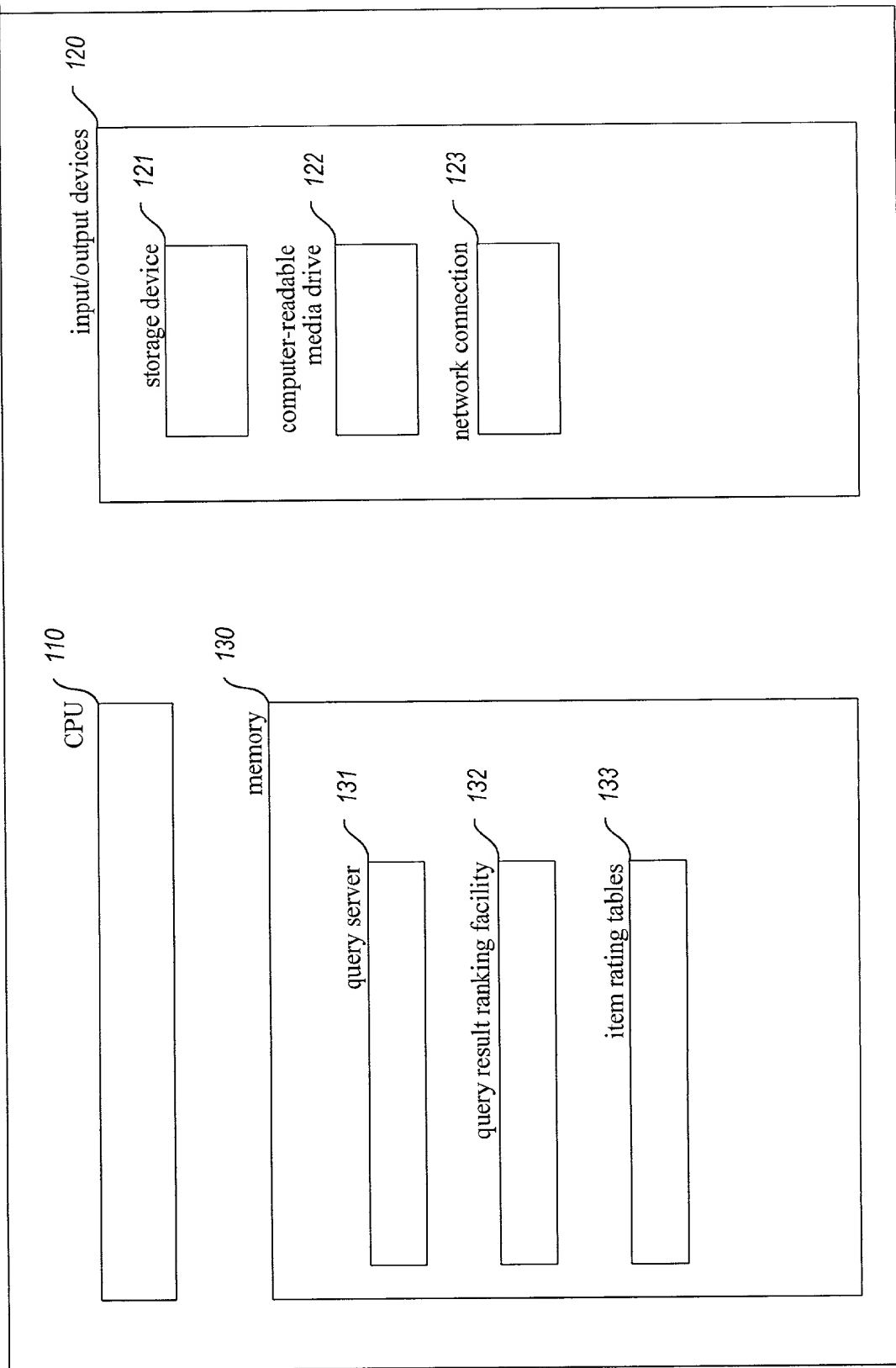


Fig. 1

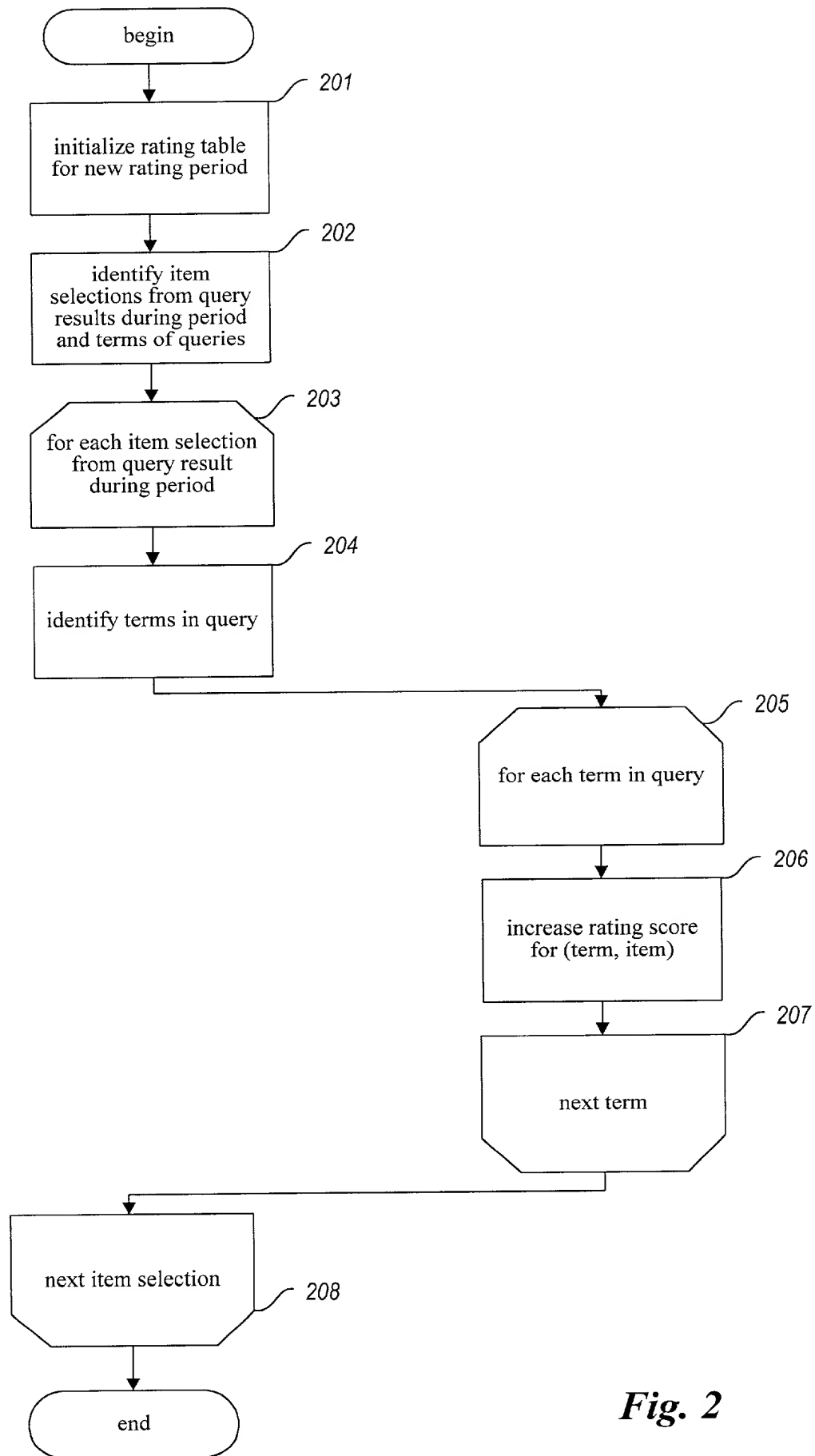


Fig. 2

item rating table			
term	item identifier	score	
	⋮		
301	dynamics	0801062272	1
302	dynamics	1883823064	22
303	dynamics	9676530409	7
	⋮		
304	human	0814403484	16
305	human	1883823064	45
306	human	6303702473	3
	⋮		

Fig. 3

400

item rating table

term	item identifier	score
	⋮	
401 dynamics	0801062272	1
402 dynamics	1883823064	23
403 dynamics	9676530409	7
	⋮	
404 human	0814403484	16
405 human	1883823064	46
406 human	6303702473	3
	⋮	

Fig. 4

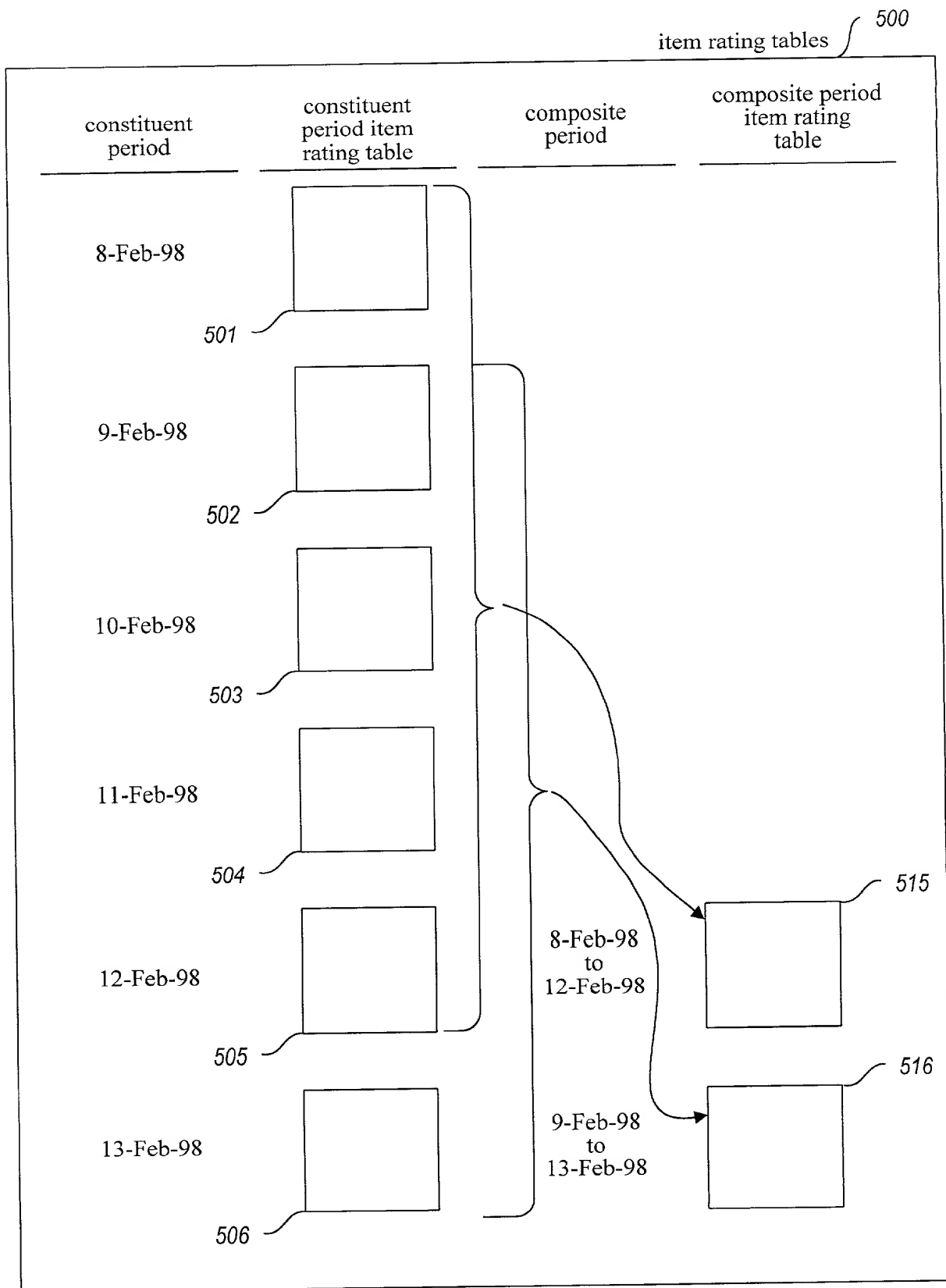


Fig. 5

600

item rating table

term	item identifier	score
	⋮	
601 dynamics	0801062272	4
602 dynamics	1883823064	116
607 dynamics	1887650024	2
603 dynamics	9676530409	45
	⋮	
604 human	0814403484	77
605 human	1883823064	211
606 human	6303702473	12
	⋮	

Fig. 6

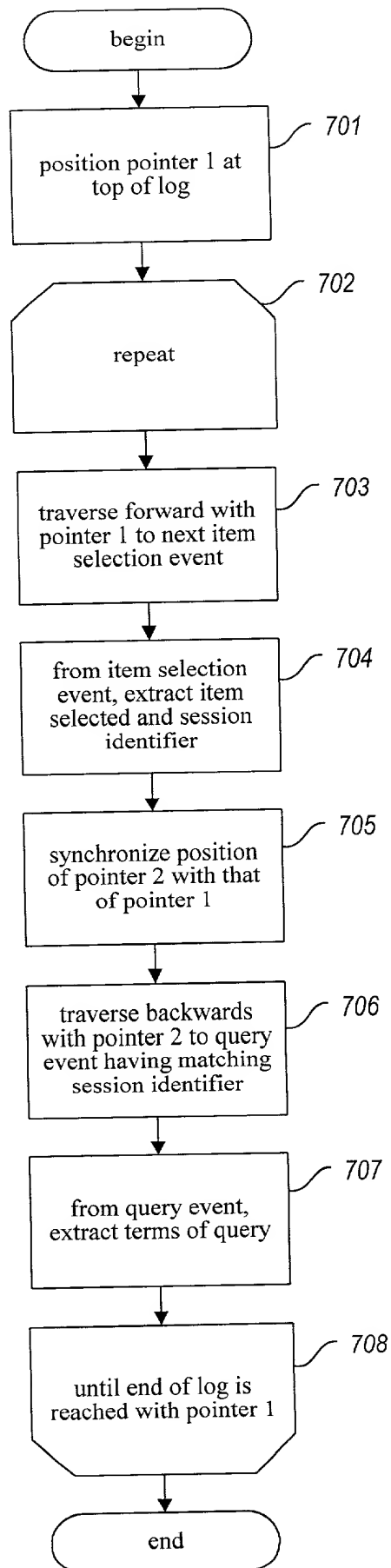


Fig. 7

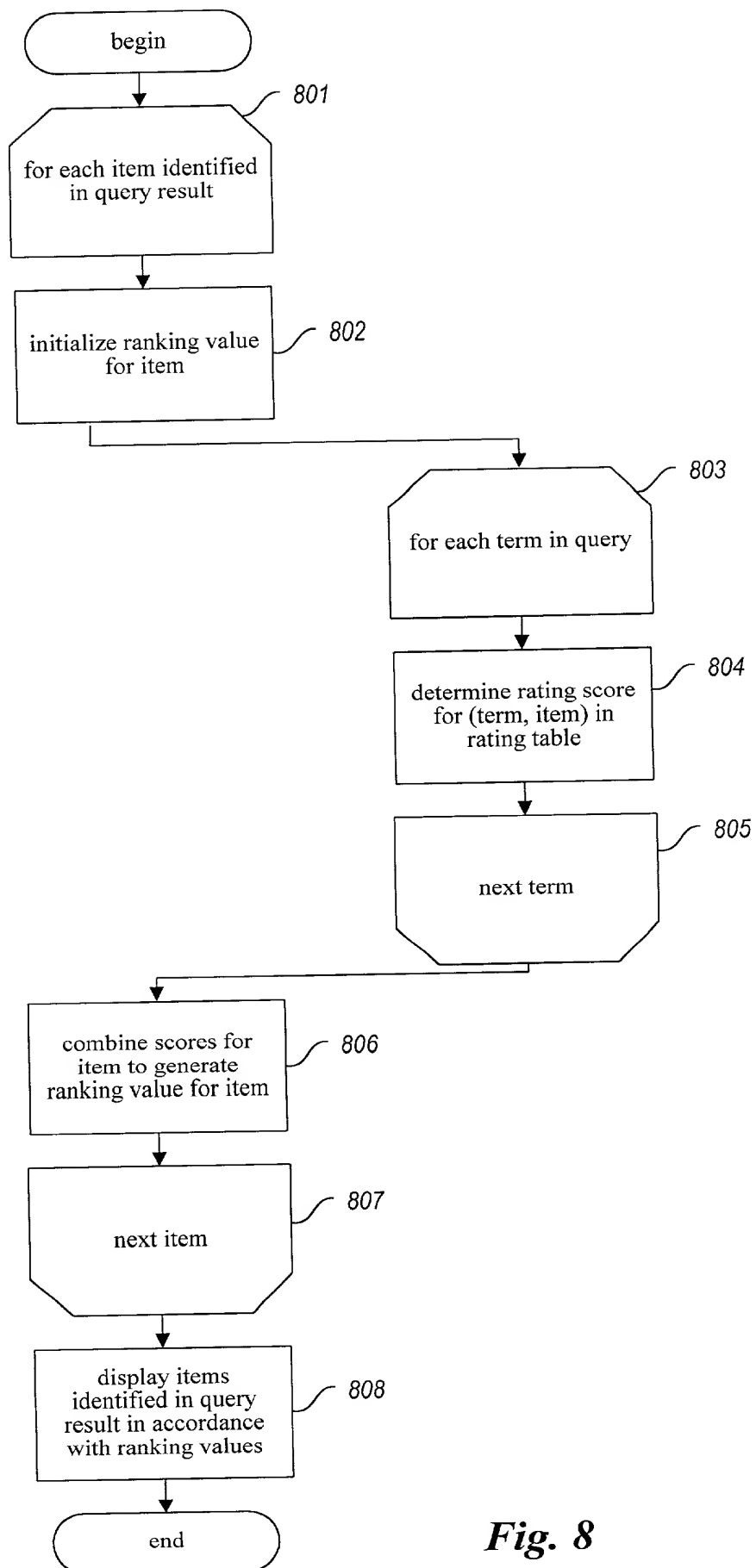


Fig. 8

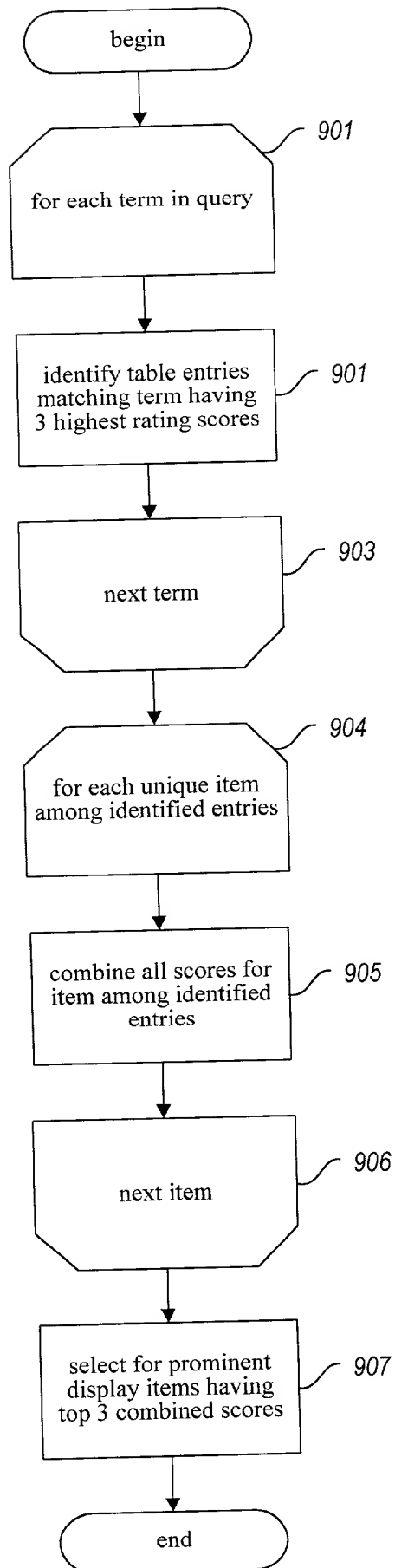


Fig. 9